High performance time series models using auto autoregressive integrated moving average

Redha Ali Al-Qazzaz, Suhad A. Yousif
Department of Computer Science, College of Science, Al-Nahrain University, Baghdad, Iraq

Article Info

Article history:

Received Jan 18, 2022
Revised May 16, 2022
Accepted May 24, 2022

Keywords:

Akaike information criterion
ARIMA
Auto ARIMA
Autocorrelation function
Mean absolute error
Partial autocorrelation function
Root mean square error

ABSTRACT

Forecasting techniques have received considerable interest from both researchers and academics because of the unique characteristics of businesses and their influence on several areas of the economy. Most academics utilize the autoregressive integrated moving average (ARIMA) approach to forecasting the future. However, researchers face challenges, such as analyzing the data and selecting the appropriate ARIMA parameters, especially with large datasets. This study investigates the use of the automatic ARIMA (Auto ARIMA) function for forecasting Brent oil prices. It demonstrates the benefits of using Auto ARIMA over ARIMA for determining the appropriate ARIMA parameters based on measures such as root mean square error (RMSE), mean absolute error (MAE), and akaike information criterion (AIC) without requiring the attention of an expert data scientist as it bypasses several steps needed for manual ARIMA. Auto ARIMA produced an RMSE of 12.5539 and an AIC of 1877.224, which are comparable to the values resulting from the manual ARIMA with the help of expert data scientists; thus, it saves analysis time and offers the best model result.

Corresponding Author:

Redha Ali Al-Qazzaz
Department of Computer Science, College of Science, Al-Nahrain University
Al Jadriyah Bridge, Baghdad 64074, Iraq
Email: redha.ali.cs2020@ced.nahrainuniv.edu.iq

1. INTRODUCTION

Today's crude oil prices have a tremendous influence on the global economy and security as oil is one of the world's primary energy sources. Because crude oil accounts for a large proportion of certain countries' exports, a rapid shift in price can have severe economic consequences, with crude oil price drops resulting in decreased economic activity [1]. Markets are becoming more competitive, especially following science and technology's rapid advancement, which forces businesses to offer a variety of high-quality items for customers while remaining cost-effective items [2]. One of the most common methods used for forecasting in various fields is the autoregressive integrated moving average (ARIMA) method, a linear time series forecasting approach used in finance, engineering, social sciences, and agriculture, among other fields [3], [4]. The ARIMA model is the result of combining autoregressive (AR) and moving average (MA) models. ARIMA have three paramerters (p, d, and q), where p represents the order of autoregressive terms, d refers to non-seasonal differences, and q denotes the order of lagged forecast errors in the forecast equation [5], [6]. ARIMA models can accurately forecast relatively steady time series data. However, they assume that future data values are proportional to present and historical data values. As a result, many real-life time series data exhibit complicated non-linear, seasonal, and non-stationary patterns that ARIMA may struggle to adequately capture [7], [8]. Several studies on crude oil time series forecasting have been conducted in recent

years [1], [9]-[12]. The problem statement to these studies, determining the three ARIMA parameters is not a simple task and required data science specialist, especially when utilizing autocorrelation and partial autocorrelation with a huge dataset. Furthermore, a unit root test such as the Dickey-Fuller test is needed to convert non-stationary data to stationary data. This conversion takes time and requires an expert to analyze the dataset, and because it may be difficult for an expert to deal with big data, the findings may occasionally be inaccurate [13]. Nevertheless, accurate time series forecasting is critical as it assists in future planning and decision-making and is the basis for greater resource.

Utilization and service levels [14]. The Akaike information criterion (AIC) is a statistical metric that can assess the relative quality of different models by comparing the goodness of fit of each model with that of other models [15]-[17]. Thus, the contributions of this paper lie in demonstrating an automate process called Auto ARIMA model, how best to utilize it, and what the advantages of Auto ARIMA are over traditional ARIMA, namely that it is faster and can directly fit the model after the data preprocessing step based on the value of the AIC. The paper is organized as follows: section 1 introduces the topic, section 2 discusses the related works, section 3 presents the methodology of time series forecasting; section 3.1 discusses traditional ARIMA, section 3.2 examines Auto ARIMA; section 4 presents the experimental results for traditional ARIMA and Auto ARIMA, and section 5 concludes the study.

2. RELATED WORK

Time series forecasts are becoming increasingly important because they are living proof and a worldwide business language [18]. Previous authors have performed several studies on the ARIMA algorithm for forecasting. For example, Siregar *et al.* [2] used SAS software and the ARIMA method to predict raw material requirements for plastic products depending on income data. MAPE was used to assess the accuracy of the predicted outcomes. The result of the forecasting for 2015 using ARIMA (3,0,2) on sales data for plastic products between 2012 and 2014 showed an increase in forecast accuracy. The research's weakness is that forecasting requires a human expert to analyze the complex data, and the model is not designed for long-term prediction; as a result, it will likely be flat or constant.

Şahinli [3], implemented an ARIMA model to predict consumer potato prices. They found the ARIMA (1,1,2) estimation to be the best model because it had the lowest criterion values, including a MAPE of 116.9075, a root mean square error (RMSE) of 201.759, and a mean absolute deviation (MAD) of 176.896.

Banerjee *et al.* [13] applied the ARIMA model (1,0,1) to predict India's approximate future stock market prices using data collected over the six years prior to the study. The results showed a root mean square (RMS) of 691.399, a mean absolute percentage error (MAPE) of 3.334, and a mean absolute error (MAE) of 506.210. The drawback of this study's approach is that it assumed that the dataset was linear, which may not have been the case. As a result, the method is rendered worthless for non-linear systems.

Fattah *et al.* [18], used the Box–Jenkins time series approach and developed an ARIMA model to estimate the finished product demand forecasting in a food factory. The most suitable model was chosen based on four criteria: AIC, standard error, Schwarz Bayesian criterion (SBC), and maximum likelihood. ARIMA (1,0,1) was chosen because it met all four preceding requirements. The results show that the model may predict future food demand. The limitation of this work is that the ARIMA model was designed to function with stationary data; hence when non-stationarity data are used, ARIMA provides low accuracy.

Ohyver and Pudjihastuti [19], proposed an ARIMA (1,1,2) model for forecasting rice prices, which was found to have good accuracy for medium-quality rice (RMSE=14.22316, AIC=4645.1), while ARIMA (2,0,2) had an RMSE of 45.53879 and an AIC of 5984.69. The drawback of the study is that ARIMA is only suitable for short-term forecasting. Therefore, it cannot be used for long-range forecasting.

Bandyopadhyay *et al.* [20], used an ARIMA (1,1,1) model to predict future gold prices based on the historical gold price data over the previous 10 years of traded values. ARIMA (1,1,1) was selected from six various model parameters because it was most effective and met all of the fit statistics criteria, whereas the other five did not. ARIMA (1,1,1) provided an RMS of 719.18, a MAPE of 3.245, and an MAE of 477.330. The challenge of the study's dataset was that under economic instability or certain government policies, it becomes impossible to record precise changes in gold prices, making the model ineffectual for forecasting in that situation. Furthermore, the approach depends on the linearity of the historical data, yet there is no proof that gold prices are linear.

Borucka [21], suggested an ARIMA (1,0,3) model based on the concept of a relationship between the values of a time series at one moment and their values at the last moments. The given model accurately predicted the number of road accidents, demonstrating its ability to forecast them. However, the study's weakness was that identification, which involves determining optimal values for function parameters, is challenging to develop for the ARIMA model.

Kumar and Vanajakshi [22], attempted to address the issue described above by offering a short-term traffic flow prediction strategy with limited input data using the seasonal ARIMA (SARIMA) model. ARIMA (2,0,0) (0,1,1) displayed an AIC of 4,218.34 which is less than that of other models. The preceding study shows that there are several obstacles and limits, but one in particular stands out: the difficulty of determining the order of the p, d, and q parameters in the ARIMA model. This is what this research paper will address.

3. RESEARCH METHOD

This section first introduces the dataset used in this study used in this research. After that, we will go through the fundamental of the forecasting strategy (ARIMA) model. Finally describes the proposed model used to predict future crude oil prices on international markets, focusing on the Auto ARIMA approach.

3.1. Dataset

There are many crude oil markets around the world. The dataset used in this paper comes from the US Energy Information Administration, which can be accessed from data.nasdaq.com. The only fields in the CSV file present dates and prices. The data comprises daily historical Brent Oil prices from 16 January 2016 through 31 December 2019. A sample of this dataset is shown in Table 1.

Table 1. Daily Brent crude oil prices

Date	Price
4-Jan-16	36.28
5-Jan-16	35.56
6-Jan-16	33.89
7-Jan-16	33.57

3.2. ARIMA

An ARIMA model combines autoregression and moving average with a difference in time series analysis. These models are used to fit data across time to improve data identification or forecast future points in the series [19]. This approach is also called the Box–Jenkins method. If the data are discovered to not be stationary, they are reduced using the differencing technique. The ARIMA approach uses three parameters: p, d, and q. The ARIMA model's p parameter indicates the number of lag periods [23], [24]. For example, if p=2 is used in the auto-regression component of the equation, two preceding periods of the time series are employed. Parameter d represents the number of differencing transformations performed to eliminate trends and/or seasonality, thereby changing the time series into a stationary one (keeping the mean and variance constant across time) [25]. This is a crucial step in preparing the data for an ARIMA model. The lag of the error component of the ARIMA model is represented by parameter q [26]. The error component is the part of the time series that cannot be explained by trend or seasonality [9]. This can also be represented as (1).

$$y_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} \quad (1)$$

Where φ_1 is the coefficient of the autoregressive model; θ is the coefficient of the moving average model; y_t is the current day, y_{t-1} is the previous day, and y_{t-2} is two days prior; and c is the constant.

The Dickey–Fuller test is used to examine time series data, stationary or not [27]. Dickey–Fuller is a statistical significance test used to determine whether null hypotheses are accepted or rejected. The null hypothesis, in this case, is that non-stationary time series data exist. If the test's p-value is greater than 0.05, the Dickey–Fuller null hypothesis will be accepted, which indicates that the data are non-stationary. Otherwise, the null hypothesis is rejected and the data are considered stationary [28]. If the time series data are non-stationary, the differencing procedure must be performed. First, the current day should be subtracted from the previous day as a differencing operation. Next, the Dickey–Fuller test should be repeated to determine whether the time series is still non-stationary. If so, then this process of running the test and performing differencing should be repeated until the time series becomes stationary [29].Furthermore, the autocorrelation function (ACF) and partial autocorrelation function (PACF) are statistical approaches for assessing how closely values in a time series are correlated, a crucial stage in ARIMA implementation. ACF and PACF plots are used to estimate the input parameters for the ARIMA model [23], [30].

The values of p and q may be determined from ACF and PACF charts using some rules. For example, if an ACF chart displays a sharp fall in autocorrelation at lag k but a smoother decrease after lag k,

then p should be zero and q should be adjusted to k (the most significant lag value). On the other hand, if the partial autocorrelation has reduced dramatically at lag k but the ACF chart indicates a smoother reduction, then q should be 0 and p should be adjusted. The general ARIMA model algorithm is presented in Algorithm 1.

```
Algorithm 1: ARIMA
Input: Crude oil dataset time series
Output: Array of forecasted values
Begin:
  #Load data of time series
  Step 1: dataset ← {x1,x2,…,xn};
  #Preprocessing Data
  Step2: All columns should be removed expect the price columns
  #Identification to determine whether or not the time series data is stationary
  Step 3: Result ←Dickey-Fuller test on the dataset;
    counter =0
      While Result < 0.05 do
        If diff == Null
        diff ← differencing(dataset); #current day - previous day
      else
        diff ← differencing(diff);
      end
        Result ←Dickey-Fuller test on the dataset;
        counter ← counter +1;
      end
  #assign d value
  Step4: d←counter
  Step5: plots ACF and PACF
  Step6: determine the order of p by observing PACF and q by observing ACF
  #Fit ARIMA model
  Step7: result_model ←ARIMA(p,d,q);
  #Forecast values on the validation set
  Step8: forecasted_array← result_model.predict(start date to end date);
  #Check the model's performance by calculating errors
  Step9: error← forecasted_array-actual_array
```

The three last stages are crucial for the time series stages. First, the predicted values depend on the variables or the other associated variables' known past values. The model is found to be appropriate in the analysis section, and the most suitable model can be utilized for future forecasts.

Finally, the future values are estimated and the model's performance is checked by calculating errors using the predictions and actual values on the validation set. The difference between the actual and predicted values is a forecast error. MAE and RMSE, described in (2) and (3), are the metrics used most often for prediction accuracy [13], [25], [31].

$$MAE = \frac{1}{n-m}\sum_{t=m+1}^{n} |y_t - \hat{y}_t| \tag{2}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n-m}\sum_{t=m+1}^{n} (y_t - \hat{y}_t)^2} \tag{3}$$

3.3. Auto ARIMA

In time series applications, many decision processes require high forecasting accuracy. ARIMA models are robust tools for time series analysis, but the model prediction using p, q, and d parameters must be analyzed [32]. Auto ARIMA saves time and perfects these parameters by iterating through the p, q, and d values. It aids in selecting the best set of these parameters and their integration into the ARIMA model [33]. Estimating the AIC can help estimate a combination of these parameters, as the best combination of p, q, and d is achieved using a lower AIC value. The auto ARIMA model helps avoid some of the steps in the ARIMA modelling technique by offering the best combination and increasing the model's performance. The following is the general auto ARIMA Model algorithm.

From comparing the two algorithms 1 and 2, we notice that the advantage of auto ARIMA over ARIMA is that auto ARIMA does not depend on manual visual observation of the ACF and PACF and dose not need an advice of an expert. Thus, it skips stages 3-5. As shown in Figure 1, from simple observation we notice that ARIMA has nine steps, while Auto ARIMA has only five steps.

```
Algorithm 2: Auto ARIMA
Input: Crude oil dataset time series
Output: Array of forecasted values
Begin:
  #Load data of time series
  Step 1: dataset ← {x1,x2,…,xn};
  #Preprocessing Data
  Step2: All columns should be removed expect the price columns
  #use Auto ARIMA model
  Step3: result_model ←Auto_ARIMA (p,d,q);
  #Forecast values on validation set
  Step4: forecasted_array← result_model.predict(start date to end date);
  #Check the model's performance by calculating errors
  Step5: error← forecasted_array-actual_array
```

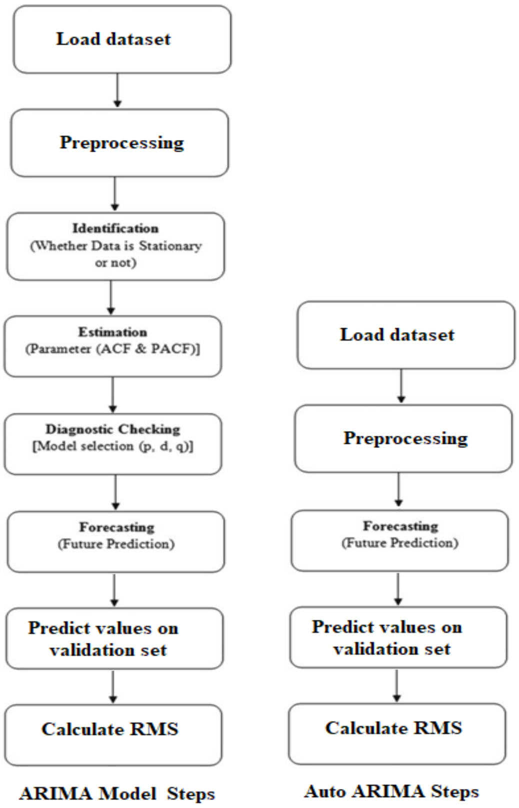

Figure 1. ARIMA and auto ARIMA steps

4. RESULTS AND DISCUSSION

This section analyzes the performance of the traditional ARIMA and auto ARIMA models using historical data from the Brent crude oil market. The original dataset was broken down into two parts: 70 percent was used as a training set for regular ARIMA and for Auto ARIMA parameter estimation, and the remaining 30 percent was used as a test set to assess the performance of the models. The results, analyses, and comparisons between models were based on many measurements such as RMSE, MAE, and AIC.

4.1. ARIMA model result

This research was carried out in a series of steps based on the Box–Jenkins approach. By relying on Algorithm 1 and after implementing the first two steps, the third step was used to determine whether the time series data were stationary or not by using the Dickey–Fuller test, because the ARIMA model only works with stationary time series data. The first Dickey–Fuller test produced a p-value of 0.4306. A p-value>0.05 signifies that the data are not stationary. If the data are non-stationary, they must be transformed using the difference operation. Step 5 utilizes ACF and PACF plots to calculate the number of orders of the p and q

parameters. The order of p can be found in the PACF plot, while q can be found in the ACF plot. In Figure 2 shows that ACF has one significant lag, and the other lag is not significant and under 0.05 (i.e., the first lag is higher than the others and more significant than the confidence area represented by the blue colour). Thus, q is inferred to be one.

Like with ACF, the same logic applies in PACF. Thus, the results of PACF are comparable to those of ACF. There is just one significant lag while the rest are not significant, so p is determined to be 1 as illustrated in Figure 3.

The difference between ACF and PACF is that instead of identifying connections between present time and lags as ACF does, PACF looks for correlations between residuals (i.e., what remains after eliminating the impacts that the previous lag) and the next lag value. Data analysis experts sometimes try multiple ARIMA models based on orders of p and q which infernces it from plots each of the ACF and PACF, for example, on the above dataset. These data analysts test ARIMA (1,0,0), then ARIMA (0,0,1), and so on, determining the percent error after each until they achieve a minimum error. As shown in Table 2.

When the three parameters (p, d, and q) were identified, the best ARIMA model was determined to be ARIMA (1,0,1). As shown in Figure 4 we notice that ARIMA (1,0,1), the green line crosses most of the points. Also ARIMA (1,0,1) has the fewest errors (MAE and RMSE), as further shown in Table 2.

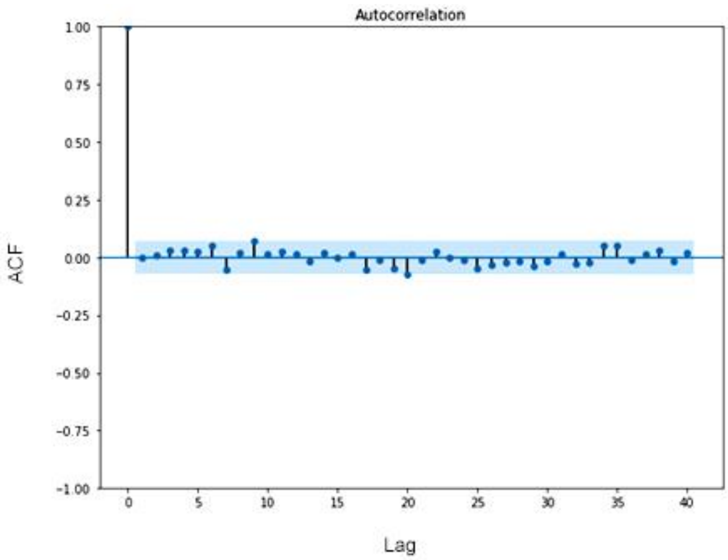

Figure 2. Autocorrelation function

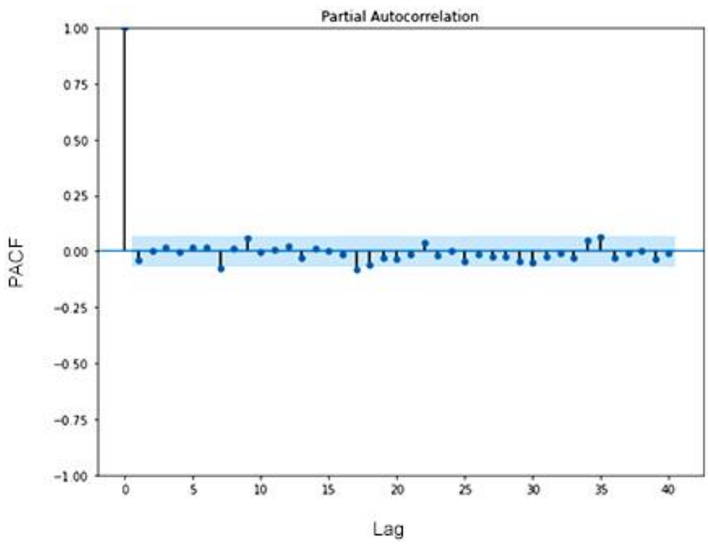

Figure 3. Partial autocorrelation

Table 2. Accuracy measures of the ARIMA model's attempt

Methods	MAE	RMSE
ARIMA (1, 0, 0)	11.4462	13.0222
ARIMA (0, 0, 1)	18.6858	19.8968
ARIMA (1, 0, 1)	10.9729	12.5539

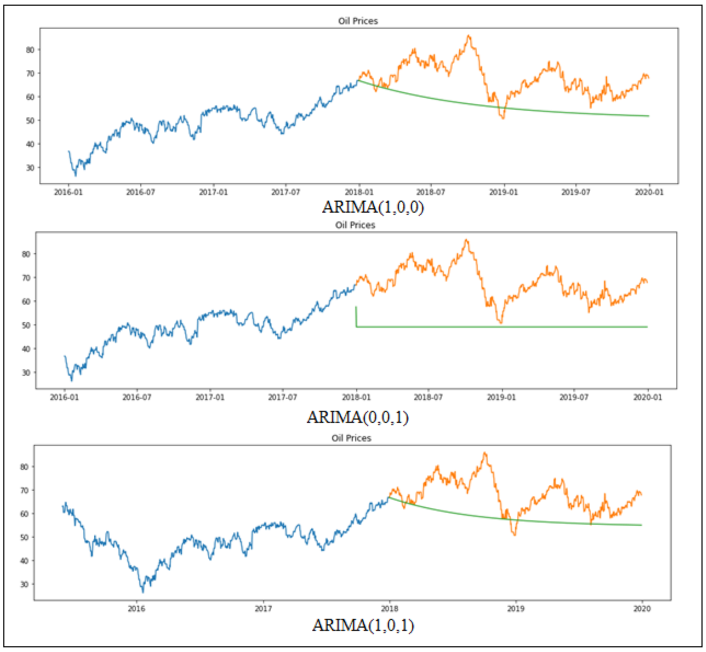

Figure 4. The green line represents the model

4.2. Auto ARIMA result

Auto ARIMA does not require the use of the Dickey–Fuller test to determine whether the data are stationary, nor does it require ACF and PACF to establish p and q. Instead, Auto ARIMA performs the three phases of the manual inspection procedure. Figure 5 shows that Auto ARIMA performed the analysis and provided parameters p, d, and q comparable to those provided by experts in manual ARIMA. It chose ARIMA (1,0,1) as the best model because it had a lower AIC of 1877.224, as illustrated in Figure 4. The performance of the models was then examined by calculating errors using the predictions and actual values from the validation set. The best fit model was estimated based on the minimal values of MAE, RMSE, and AIC, as shown in Table 3. Therefore, the ARIMA model (1,0,1) was most suitable for this dataset.

```
Performing stepwise search to minimize aic
 ARIMA(0,0,0)(0,0,0)[0]             : AIC=7773.871, Time=0.02 sec
 ARIMA(1,0,0)(0,0,0)[0]             : AIC=inf, Time=0.17 sec
 ARIMA(0,0,1)(0,0,0)[0]             : AIC=inf, Time=0.18 sec
 ARIMA(1,0,1)(0,0,0)[0]             : AIC=1879.000, Time=0.33 sec
 ARIMA(1,0,1)(0,0,0)[0] intercept   : AIC=1877.224, Time=0.40 sec
 ARIMA(0,0,1)(0,0,0)[0] intercept   : AIC=4232.578, Time=0.19 sec
 ARIMA(1,0,0)(0,0,0)[0] intercept   : AIC=inf, Time=0.19 sec
 ARIMA(0,0,0)(0,0,0)[0] intercept   : AIC=5121.293, Time=0.03 sec

Best model:  ARIMA(1,0,1)(0,0,0)[0] intercept
Total fit time: 1.513 seconds
```

Figure 5. Trace of auto ARIMA

From all the phases of manual ARIMA, steps 3, 4, and 6 were shortened while the results remained the same. This eliminates the need for professionals when analyzing the data and selecting parameter values.

Depending only on manual visual inspections and following expert-recommended guidelines may predict a small univariate time series dataset, such as our dataset above. However, in the case of enormous datasets, e.g., IoT sensors, a manual visual inspection would not be possible or would take a tremendous amount of time. One solution to this is to employ Auto ARIMA, which allows for the automatic selection of ARIMA parameters to choose the most optimal model.

Table 3. The accuracy of the forecasting model

Methods	MAE	RMSE	AIC
ARIMA (1,0,1)	10.9729	12.5539	1877.224
ARIMA (1,0,0)	11.4462	13.0222	INF
ARIMA (0,0,1)	18.6858	19.8968	4232.578

5. CONCLUSION

This study revealed that the most challenging stage of the traditional ARIMA model is identification, the process of determining optimal values for function parameters that requires the participation of experts. Therefore, auto ARIMA is recommended for beginners data scientist in forecasting in this study because it abbreviates the most difficult and unpleasant stage in the dataset analysis. It select the best combination of parameters by using the AIC to compare models and choose the best one. Auto ARIMA chose the best model with the lowest AIC. This saves a substantial amount of time and eliminates the need to understand the statistics and theory underlying the model selection. Moreover, this strategy reduces the risk of human error and the possibility for errors produced by incorrect interpretation of the results.

REFERENCES

[1] J. Guo, "Oil price forecast using deep learning and ARIMA," in *2019 International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI) , 2019*, pp. 241-247, doi: 10.1109/MLBDBI48998.2019.00054.

[2] B. Siregar, E. B. Nababan, A. Yap, and U. Andayani, "Forecasting of raw material needed for plastic products based in income data using ARIMA method," in *2017 5th International Conference on Electrical, Electronics and Information Engineering (ICEEIE)*, 2017, pp. 135-139, doi: 10.1109/ICEEIE.2017.8328777.

[3] M. A. Şahinli, "Potato price forecasting with holt-winters and ARIMA methods: A case study," *American Journal of Potato Research,* vol. 97, no. 4, pp. 336-346, 2020, doi: 10.1007/s12230-020-09788-y.

[4] M. A. A. Amin and M. A. Hoque, "Comparison of ARIMA and SVM for short-term load forecasting," in *2019 9th Annual Information Technology, Electromechanical Engineering and Microelectronics Conference (IEMECON)*, 2019, pp. 1-6, doi: 10.1109/IEMECONX.2019.8877077.

[5] G. Aradhye, A. C. S. Rao, and M. D. M. Mohammed, "A novel hybrid approach for time series data forecasting using moving average filter and ARIMA-SVM," in *Advances in Intelligent Systems and Computing*, 2019, pp. 369-381, doi: 10.1007/978-981-13-1498-8_33.

[6] V. Hryhorkiv, L. Buiak, A. Verstiak, M. Hryhorkiv, O. Verstiak, and K. Tokarieva, "Forecasting financial time sesries using combined ARIMA-ANN Algorithm," in *2020 10th International Conference on Advanced Computer Information Technologies (ACIT)*, 2020, pp. 455-458, doi: 10.1109/ACIT49673.2020.9208859.

[7] Ü. Ç. Büyükşahin and Ş. Ertekin, "Improving forecasting accuracy of time series data using a new ARIMA-ANN hybrid method and empirical mode decomposition," *Neurocomputing,* vol. 361, pp. 151-163, 2019, doi: 10.1016/j.neucom.2019.05.099.

[8] S. S. Namini, N. Tavakoli, and A. S. Namin, "A comparison of ARIMA and LSTM in forecasting time series," in *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 2018, pp. 1394-1401, doi: 10.1109/ICMLA.2018.00227.

[9] B. J. Ayeni and R. Pilat, "Crude oil reserve estimation: an application of the autoregressive integrated moving average (ARIMA) model," *Journal of Petroleum Science and Engineering*, vol. 8, no. 1, pp. 13-28,doi: 10.1016/0920-4105(92)90041-X.

[10] Y. Chen, K. He, and G. K. F. Tso, "Forecasting crude oil prices: a deep learning based model," *Procedia Computer Science*, vol. 122, pp. 300-307, 2017, doi: 10.1016/j.procs.2017.11.373.

[11] E. K. Mensah, "Box-Jenkins modelling and forecasting of Brent crude oil price," Munich Personal RePEc Archive, 2015. [Online]. Available: https://mpra.ub.uni-muenchen.de/67748/14/MPRA_paper_67748.pdf

[12] C. l. Zhao and B. Wang, "Forecasting crude oil price with an autoregressive integrated moving average (ARIMA) model," in *Advances in Intelligent Systems and Computing*, 2014, pp. 275-286, doi: 10.1007/978-3-642-38667-1_27.

[13] D. Banerjee, "Forecasting of Indian stock market using time-series ARIMA model," in *2014 2nd International Conference on Business and Information Management (ICBIM)*, 2014, pp. 131-135, doi: 10.1109/ICBIM.2014.6970973.

[14] Y. Deng, H. Fan, and S. Wu, "A hybrid ARIMA-LSTM model optimized by BP in the forecast of outpatient visits," *Journal of Ambient Intelligence and Humanized Computing,* 2020, doi: 10.1007/s12652-020-02602-x.

[15] T. Rutkowski, K. Łapa, M. Jaworski, "On explainable flexible fuzzy recommender and its performance evaluation using the akaike information criterion," in *Neural Information Processing. ICONIP 2019. Communications in Computer and Information Science,* 2019, pp. 717-724, doi: 10.1007/978-3-030-36808-1_78.

[16] S. Atique, S. Noureen, V. Roy, V. Subburaj, S. Bayne, and J. Macfie, "Forecasting of total daily solar energy generation using ARIMA: A case study," in *2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC)*, 2019, pp. 0114-0119, doi: 10.1109/CCWC.2019.8666481.

[17] N. T. Tran and L. V. Dai, "Grid search of exponential smoothing method: a case study of Ho Chi Minh City load demand," *Indonesian Journal of Electrical Engineering and Computer Science,* vol. 19, no. 3, pp. 1121-1130, 2020, doi: 10.11591/ijeecs.v19.i3.pp1121-1130.

[18] J. Fattah, L. Ezzine, Z. Aman, H. El Moussami, and A. Lachhab, "Forecasting of demand using ARIMA model," *International Journal of Engineering Business Management*, vol. 10, 2018, doi: 10.1177/1847979018808673.

[19] M. Ohyver and H. Pudjihastuti, "Arima model for forecasting the price of medium quality rice to anticipate price fluctuations," *Procedia Computer Science*, vol. 135, pp. 707-711, 2018, doi: 10.1016/j.procs.2018.08.215.

[20] B. Guha and G. Bandyopadhyay, "Gold price forecasting using ARIMA model," *Journal of Advanced Management Science,* vol. 4, no. 2, pp. 117-121, 2016, doi: 10.12720/joams.4.2.117-121.

[21] A. Borucka, "Risk analysis of accidents in Poland based on ARIMA Model," in *Proceedings of 22ndInternational Scientific Conference*, 2018, pp. 162-166.

[22] S. V. Kumar and L. Vanajakshi, "Short-term traffic flow prediction using seasonal ARIMA model with limited input data," *European Transport Research Review,* vol. 7, no. 3, p. 21, 2015, doi: 10.1007/s12544-015-0170-8.

[23] R. Madan and P. S. Mangipudi, "Predicting computer network traffic: a time series forecasting approach using DWT, ARIMA and RNN," in *2018 Eleventh International Conference on Contemporary Computing (IC3)*, 2018, pp. 1-5, doi: 10.1109/IC3.2018.8530608.

[24] I. Mado, A. Soeprijanto, and S. Suhartono, "Applying of double seasonal ARIMA model for electrical power demand forecasting at PT. PLN Gresik Indonesia," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 8, no. 6, pp. 4892-4901, 2018, doi: 10.11591/ijece.v8i6.pp4892-4901.

[25] O. Yakubu and N. Babu, "Electricity consumption forecasting using DFT decomposition based hybrid ARIMA-DLSTM model," *Indonesian Journal of Electrical Engineering and Computer Science,* vol. 24, no. 2, pp. 1107-1120, 2021, doi: 10.11591/ijeecs.v24.i2.pp1107-1120.

[26] B. A. Hussain, "Simulating an evolution strategy to forecast time series ARMA (1,1) model," *Journal of Al-Nahrain University*, vol. 10, no. 1, pp. 167-175, 2007, doi: 10.22401/JNUS.10.1.29.

[27] A. P. Behera, M. K. Gaurisaria, S. S. Rautaray and M. Pandey, "Predicting future call volume using ARIMA Models," in *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 2021, pp. 1351-1354, doi: 10.1109/ICICCS51141.2021.9432314.

[28] F. X. Diebold and G. D. Rudebusch, "On the power of Dickey-Fuller tests against fractional alternatives," *Economics Letters*, vol. 35, no. 2, pp. 155-160, 1991, doi: 10.1016/0165-1765(91)90163-F.

[29] M. T. Cao, T. T. Pham, T. C. Kuo, D. M. Bui, H. V. Nguyen, and T. H. Nguyen, "Short-term load forecasting enhanced with statistical data-filtering method," in *2020 IEEE International Conference on Power Electronics, Smart Grid and Renewable Energy (PESGRE2020)*, 2020, pp. 1-8, doi: 10.1109/PESGRE45664.2020.9070497.

[30] M. A. Masngut, S. Ismail, A. Mustapha, and S. M. Yasin, "Comparison of daily rainfall forecasting using multilayer perceptron neural network model," *IAES International Journal of Artificial Intelligence (IJ-AI),* vol. 9, no. 3, pp.456-463, 2020, doi: 10.11591/ijai.v9.i3.pp456-463.

[31] P. Areekul, T. Senjyu, H. Toyama and A. Yona, "Notice of violation of IEEE publication principles: a hybrid ARIMA and neural network model for short-term price forecasting in deregulated market," *IEEE Transactions on Power Systems*, vol. 25, no. 1, pp. 524-530, Feb. 2010, doi: 10.1109/TPWRS.2009.2036488.

[32] H. Bousqaoui, I. Slimani, and S. Achchab, "Comparative analysis of short-term demand predicting models using ARIMA and deep learning," *International Journal of Electrical and Computer Engineering (IJECE),* vol. 11, no. 4, pp. 3319-3328, 2021, doi: 10.11591/ijece.v11i4.pp3319-3328.

[33] H. Alabdulrazzaq, M. N. Alenezi, Y. Rawajfih, B. A. Alghannam, A. A. Al-Hassan, and F. S. Al-Anzi, "On the accuracy of ARIMA based prediction of COVID-19 spread," *Results in Physics,* vol. 27, p. 104509, 2021, doi: 10.1016/j.rinp.2021.104509.

BIOGRAPHIES OF AUTHORS

Redha Ali Al-Qazzaz is a Master's Student at Al-Nahrain University/College of Science/Computer science department. He received his BSc from Al-Nahrain University in 2011, Redha currently works as a developer in the Iraqi Ministry of Oil/State Oil Marketing Organization (SOMO). His special area is in Machine learning, DBA Databases, Full Stack Developer, and Cloud Computing. He can be contacted at email: redha.ali.cs2020@ced.nahrainuniv.edu.iq.

Suhad A. Yousif She is an assistant professor at Al-Nahrain University/College of science. She is head of the Computer science department. She received her BSc from Al-Nahrain University in 1994, the M.Sc in Computer Science department/Baghdad university in 2005, and Ph.D. degrees from Mathematics and Computer Science Department in Beirut Arab University/Lebanon in 2015. Dr. Suhad supervises M.Sc. theses concerning cloud computing, big data analysis, text classification (natural language processing), and classification of Ensemble machine learning, Automated machine learning, and forecasting prediction. She also leads and teaches different subjects at both B.Sc. and M.Sc. Levels in computer science. In addition, she is on a scientific committee at some conferences and a reviewer in several conferences and Journals. Her particular area of research is Big data, Data Science, Machine learning, and deep learning. She can be contacted at email: redha.ali.cs2020@ced.nahrainuniv.edu.iq.